



HAL
open science

Preference Heterogeneity in Experiments: Comparing the Field and Laboratory

Steffen Andersen, Glenn W. Harrison, Morten Igel Lau, E. Elisabet Rutström

► **To cite this version:**

Steffen Andersen, Glenn W. Harrison, Morten Igel Lau, E. Elisabet Rutström. Preference Heterogeneity in Experiments: Comparing the Field and Laboratory. *Journal of Economic Behavior and Organization*, 2009, 73 (2), pp.209. 10.1016/j.jebo.2009.09.006 . hal-00736951

HAL Id: hal-00736951

<https://hal.science/hal-00736951>

Submitted on 1 Oct 2012

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Accepted Manuscript

Title: Preference Heterogeneity in Experiments: Comparing the Field and Laboratory

Authors: Steffen Andersen, Glenn W. Harrison, Morten Igel
Lau, E. Elisabet Rutström



PII: S0167-2681(09)00231-5
DOI: doi:10.1016/j.jebo.2009.09.006
Reference: JEBO 2452

To appear in: *Journal of Economic Behavior & Organization*

Received date: 3-8-2006
Revised date: 23-9-2009
Accepted date: 23-9-2009

Please cite this article as: Andersen, S., Harrison, G.W., Lau, M.I., Rutström, E.E., Preference Heterogeneity in Experiments: Comparing the Field and Laboratory, *Journal of Economic Behavior and Organization* (2008), doi:10.1016/j.jebo.2009.09.006

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Preference Heterogeneity in Experiments: Comparing the Field and Laboratory

by

Steffen Andersen, Glenn W. Harrison, Morten Igel Lau and E. Elisabet Rutström[†]

September 2009

Abstract. Do laboratory experiments provide a reliable basis for measuring field preferences? Economists recognize that preferences can differ across individuals, but only a few attempts have been made to elicit individual preferences for representative samples of a population in a particular geographical area, region or country. Our primary objective is to compare directly estimates of preferences elicited from a convenience sample and the estimates for the wider population from which that sample is drawn. We examine the strengths and weaknesses of laboratory and field experiments to detect differences in preferences over risk and time that are associated with standard, observable characteristics of the individual.

[†] Department of Economics and Centre for Economic and Business Research, Copenhagen Business School, Copenhagen, Denmark (Andersen); Department of Risk Management & Insurance and CEAR, Robinson College of Business, Georgia State University, USA (Harrison); Department of Economics, College of Business Administration, University of Central Florida, USA (Rutström) and Department of Economics, Newcastle University Business School, Newcastle University, England (Lau). E-mail: SA.ECO@CBS.DK, GHARRISON@GSU.EDU, M.I.LAU@DURHAM.AC.UK, and ERUTSTROM@BUS.UCF.EDU. We thank the U.S. National Science Foundation for research support under grants NSF/DRU 0527675 and NSF/SES 0616746, and the Danish Social Science Research Council for research support under projects #24-02-0124 and #275-08-0289. The coeditor Catherine Eckel and two referees have provided many helpful comments. Supporting data and statistical code are stored in the *ExLab* Digital Archive at <http://exlab.bus.ucf.edu>.

Do laboratory experiments provide a reliable basis for measuring field preferences? There have been tremendous advances in the application of experimental methods in laboratory settings and widespread acceptance of the results. Laboratory results on the elicitation of preferences have significantly influenced the development of game theory, auction theory, utility theory, and environmental damage assessment.¹ Implicit in these applications is the presumption that laboratory results are reliable assessments of general results, even though we know that the typical subject pool is different from the population to which they are being applied. Economists recognize that preferences can differ across individuals, but only a few attempts have been made to elicit individual preferences for representative samples of a population in a particular geographical area, region or country.² We compare estimates of preferences elicited from a sample of university students in Copenhagen and estimates from a sample of the adult population in Denmark.

We focus on the strengths and weaknesses of field and laboratory experiments to detect differences in preferences in two separate dimensions: aversion to risk and time delay. We use experimental tasks with real monetary rewards to elicit individual risk attitudes and discount rates. These valuation tasks build on the risk aversion experiments of Holt and Laury (2002) and Harrison et al. (2005b) and the discount rate experiments of Coller and Williams (1999) and Harrison et al. (2002b). Our experimental design is implemented in the field in Denmark to obtain a representative sample of the adult population that offers a wider range of individual characteristics than usually found in subject pools recruited at universities.³ We also conducted a similar set of laboratory

¹ Recent surveys include Camerer (2003), Kagel and Levin (2002), Starmer (2000) and Harrison (2006), respectively.

² Harrison et al. (2002b) elicit individual discount rates for a representative sample of the adult Danish population and find evidence of significant preference heterogeneity across socio-demographic variables. This is the first attempt to elicit individual preferences of a population in a country using controlled experiments with monetary rewards. Eckel et al. (2005) conduct a field study of time and risk preferences. Their subjects are recruited from low income neighborhoods in Montreal, and they are given 64 “compensated” questions, one of which is chosen at random for payment. Dohmen et al. (2005) elicit individual risk attitudes and combine hypothetical surveys with experiments that involve monetary incentives. A total of 450 subjects participated in the experiment, and they were recruited from 179 randomly chosen voting districts in Germany.

³ Our field experiments are “artefactual field experiments” in the terminology of Harrison and List (2004), who also survey the literature. This type of experiment involves the use of laboratory tasks and procedures in the field, using a subject pool that is more representative of the target population about which one wants to make an inference. Field experiments include other types of experiments, in which one uses procedures or information sets that occur naturally in the field. One virtue of artefactual field experiments is that one has the same control over procedures and information that one has in conventional laboratory experiments, making comparisons of results, such as we do here, easier. One

experiments in Copenhagen to test for preference heterogeneity across the two samples.

Field and laboratory experiments are *complementary* tools in the evaluation of risk and time preferences. The laboratory provides a cost-effective proving ground for experimental designs and basic treatments, but with the concern that the sample and task are not representative of those faced in the field. The use of field experiments allows examination of the applicability of findings from the laboratory to the population of interest for policy evaluation, so the two types of experiments are best viewed as complementary (Harrison and List (2004)). The simplest type of field experiment, used here, uses laboratory instruments and procedures with a target population drawn at a national level.

Risk attitudes for the population are critical inputs into cost-benefit analyses of policy whenever there is uncertainty about the size of the costs or benefits. Our interest in eliciting risk attitudes from a broader sample of the Danish adult population derived from quantitative welfare simulations of policy reforms, such as tax uniformity (Harrison et al. 2002a). Welfare estimations entail uncertain impacts due to uncertainty about parameters of the underlying model (Harrison and Vinod 1992). When we present policy makers with uncertain estimates of the purported economic impacts of these reforms, we therefore have to acknowledge that we are implicitly asking the policy maker to decide on a “policy lottery” for Danes. Armed with estimates of the risk attitudes of the Danish population, we can evaluate the certainty-equivalent welfare impact across Danes. The same general motivation applies for the elicitation of discount rates. Most policies have a time dimension in the sense that expected costs and benefits vary over some time horizon. Hence any present value evaluation of welfare effects has to consider the time preferences of those affected by the policy (e.g., Lau 2000).

The United States Office of Management (2006) has proposed guidelines for the conduct

important extension of our approach would be to examine if field *surveys* generated reliable estimates of risk aversion and discount rates. Those surveys have the advantage that they can be sent out to much larger samples than we could afford to visit using artefactual field experiments. They have the obvious disadvantage that they involve hypothetical tasks, but there is now a substantial literature on the complementary use of (laboratory) experiments and surveys in which the laboratory results are used to “statistically calibrate” the surveys for the effects of hypothetical bias (e.g., Blackburn et al. (1994)).

and use of surveys by federal government agencies. One recurring theme is the importance of the sample being representative of the target population (Standard 1.3). The same logic would be expected to extend to the use of experiments by federal government agencies. Thus the comparability of preferences elicited from the laboratory and field samples is of direct relevance for the policy applicability of results from experiments. In addition, these guidelines stress the need for full pre-tests of all field instruments (Standard 1.4), which we interpret as pointing to another sense in which laboratory experiments complement field experiments.

We find similar preferences on average across the field and laboratory samples. However, we are able to detect a much richer preference heterogeneity when we sample from the field as compared to convenience samples drawn from college students in our laboratory experiments. We say “detect” because we do not claim that the heterogeneity is missing in the sample used in the laboratory, just that the range of variation in characteristics is smaller. The convenience samples are more homogeneous with respect to many observable characteristics such as age and education level. Since we have the benefit of comparable laboratory and field experiments, we can claim that there is preference heterogeneity in the field, so we know that we are missing something in the laboratory. We also find significant interaction effects between the treatment effects and the individual characteristics in the field data. These effects are observed for both age and sex for our risk attitude tests, and we also find significant interaction effects between sex and framing for our discount rate tests.

1. Treatment Effects and Heterogeneity

The goal of any evaluation method for treatment effects is to construct the proper counterfactual, and economists have spent years examining approaches to this problem. Harrison and List (2004; p.1014ff.) review five alternative methods of constructing the counterfactual: controlled experiments, natural experiments, propensity score matching (PSM), instrumental variables estimation, and structural approaches. We review the first and third here to be able to state the main hypothesis of our study.

Define y_1 as the outcome with treatment, y_0 as the outcome without treatment, and let $T = 1$ when treated and $T = 0$ when not treated.⁴ The treatment effect for unit i can then be measured as $\tau_i = y_{i1} - y_{i0}$. The major problem, however, is one of a missing counterfactual: τ_i is unknown. If we could observe the outcome for an untreated observation had it been treated, then there is no evaluation problem.⁵

“Controlled” experiments, which include laboratory experiments and field experiments, represent the most convincing method of creating the counterfactual since they directly construct a control group via randomization.⁶ In this case, the population average treatment effect is given by $\tau = y^*_1 - y^*_0$, where y^*_1 and y^*_0 are the treated and non-treated average outcomes after the treatment. The assumption is that the population from which the samples are being randomized is representative of the target population for inferences about treatment effects.

One alternative method of assessing the impact of the treatment is the method of propensity score matching (PSM) developed in Rosenbaum and Rubin (1983). This method has been used extensively in the debate over experimental and non-experimental evaluation of treatment effects initiated by Lalonde (1986; see also Dehejia and Wahba 1999 and 2002 and Smith and Todd 2005). The goal of PSM is to make non-experimental data “look like” experimental data. The intuition behind PSM is that if the researcher can select observable factors so that any two individuals with the same value for these factors will display homogenous responses to the treatment, then the treatment effect can be measured without bias. In effect, one can use statistical methods to identify which two individuals are more homogeneous for the purposes of measuring the treatment effect. More formally, the solution advocated is to find a vector of covariates, Z , such that $y_1 y_0 \perp T \mid Z$ and

⁴ We simplify by considering a binary treatment, but the logic generalizes easily to multiple treatment levels and continuous treatments. Obvious examples from outside economics include dosage levels or stress levels. In economics, one might have some measure of risk aversion or “other regarding preferences” as a continuous treatment.

⁵ A within-sample design where each subject is presented to all treatments does not solve the problem of a missing counterfactual. In that case one would have to control for order effects, and a between-sample design is needed to control and test for those. One prominent example in the case of risk aversion using the Holt and Laury (2002) instrument is provided by Harrison et al. (2005b). They show that order effects do have a statistically significant effect on risk attitudes.

⁶ Experiments are often run in which the control is provided by theory, and the objective is to assess how well theory matches behavior. This would seem to rule out a role for randomization, until one recognizes that some implicit or explicit error structure is required in order to test theories meaningfully (Ballinger and Wilcox 1997).

$\text{pr}(T=1 \mid Z) \in (0,1)$, where \perp denotes independence.⁷

We can contrast the manner in which controlled experiments and statistical methods such as PSM ensure reliable measurement of treatment effects. In the former there is an *ex ante* randomization to treatment, and in the latter there is an *ex post* conditioning of the sample to reduce the effects of heterogeneity. The problem with laboratory experiments, however, is that they reflect a very special population when conducted on university students. Such experiments, using randomization to treatment within that special population, can make reliable measurements of treatment effects for that special population.⁸ However, it does not follow that they can make reliable measurements of treatment effects for the broader population. A similar argument can be made with respect to our artefactual field experiments in Denmark. It does not follow that they generate reliable measurements of treatments effects for populations in other countries.

2. Valuation Tasks

A. Risk Aversion

We employ a simple experimental measure for risk aversion introduced by Holt and Laury (2002) and extended by Harrison et al. (2005). Each subject is presented with a choice between two lotteries, which we can call A or B. Table 1 illustrates the basic payoff matrix presented to subjects using lottery prizes from Holt and Laury (2002). The first row shows that lottery A offered a 10% chance of receiving \$2 and a 90% chance of receiving \$1.60. The expected value of this lottery, EV^A , is shown in the third-last column as \$1.64, although the EV columns were not presented to subjects. Similarly, lottery B in the first row has chances of payoffs of \$3.85 and \$0.10, for an expected value of \$0.48. Thus the two lotteries have a relatively large difference in expected values, in this case

⁷ If one is interested in estimating the average treatment effect, only the weaker condition $E[y_0 \mid T=1, Z] = E[y_0 \mid T=0, Z] = E[y_0 \mid Z]$ is required. This assumption is called the “conditional independence assumption,” and intuitively means that given Z , the non-treated outcomes are what the treated outcomes would have been had they not been treated. Treated, or likewise, that selection occurs only on observable variables. Note that the dimensionality of the problem, as measured by Z , may limit the use of matching. A more feasible alternative is to match on a function of Z . Rosenbaum and Rubin (1983 and 1984) showed that matching on $p(Z)$ instead of Z is valid. This is usually carried out on the “propensity” to get treated $p(Z)$, or the propensity score, which in turn is often implemented by a simple probit or logit model with T as the dependent variable.

⁸ Experimental economists are notoriously casual about the sample sizes needed for randomization to work its statistical magic and virtually never undertake power calculations before going into the lab. We conjecture that many claims in the experimental literature have extremely lower statistical power.

\$1.16. As one proceeds down the matrix, the expected value of both lotteries increases and the expected value of lottery B eventually exceeds the expected value of lottery A.

The subject chooses A or B in each row. The logic behind this test for risk aversion is that only risk-loving subjects would take lottery B in the first row, and only risk-averse subjects would take lottery A in the second last row. A risk neutral subject should switch from choosing A to B when the EV of each is about the same, so a risk-neutral subject would choose A for the first four rows and B thereafter. In addition to the A/B choice on each row there is also an option to express indifference, which is not shown in Table 1.

These data may be analyzed using a constant relative risk aversion (CRRA) characterization of utility, employing an interval regression model.⁹ The CRRA utility of each lottery prize y is defined as $U(y) = (y^{1-r})/(1-r)$, where r is the CRRA coefficient.¹⁰ The dependent variable in the interval regression model is the CRRA interval that subjects implicitly choose when they switch from lottery A to lottery B. For each row of Table 1 one can calculate the implied bounds on the CRRA coefficient, and these intervals are shown in the final column of Table 1. Thus, for example, a subject who made 5 safe choices and then switched to the risky alternatives would have revealed a CRRA interval between 0.14 and 0.41, a subject who made 7 safe choices would have revealed a CRRA interval between 0.68 and 0.97, and so on.¹¹

We undertake four separate risk aversion tasks with each subject in the field and laboratory experiments. The payoffs were in Danish Kroner (DKK), and each task had 4 different prizes, selected so that all 16 prizes span the range of income over which we seek to estimate risk aversion. The four sets of prizes are as follows (in Danish kroner), with the two prizes for lottery A listed first

⁹ Holt and Laury (2002) introduce the use of the flexible Expo-Power (EP) utility function, originally developed by Saha (1993), for the analysis of risk aversion in experiments such as these. EP allows varying RRA over the prize income subjects faced in an experiment, thereby generalizing CRRA. Harrison, Lau and Rutström (2007) find that CRRA cannot be rejected as an overall characterization of behavior in the field experiments, and the estimates indicate that there is no significant evidence against CRRA for any of the individual characteristics controlled for.

¹⁰ With this parameterization, $r = 0$ denotes risk neutral behavior, $r > 0$ denotes risk aversion, and $r < 0$ denotes risk loving. When $r = 1$, $U(y) = \ln(y)$.

¹¹ Following Rabin (2000), there are some specifications of expected utility theory for which a finding of risk aversion at these levels of income is incoherent. This argument does not apply if expected utility is defined over income earned during the experiment rather than over terminal lifetime wealth. Such specifications are standard in experimental economics, as well as large areas of economic theory such as the analysis of auctions and contracts. Cox and Sadiraj (2006), Harrison et al. (2007; Appendix A) and Andersen et al. (2008a) review these methodological issues in further detail.

and the two prizes for lottery B listed next: (A1: 2000, 1600; B1: 3850, 100), (A2: 2250, 1500; B2: 4000, 500), (A3: 2000, 1750; B3: 4000, 150), and (A4: 2500, 1000; B4: 4500, 50). At the time of the field experiments, the exchange rate was approximately 6.55 kroner per U.S. dollar, so the prizes range from approximately 7.65 to 687 dollars.

To allow more refined elicitation of the true preferences, yet retain the transparency of the incentives of the basic multiple price list, we used a computerized variant on the MPL format which we call an Iterative MPL (iMPL). The basic MPL is the standard format in which the subject sees a fixed array of paired options and chooses one for each row. It allows subjects to switch back and forth as they like and has already been used in many experiments. The iMPL format extends this by first asking the subject simply to choose the row at which he wants to first switch from option A to option B, assuming monotonicity of the underlying preferences to fill out the remaining choices automatically. The second extension of the MPL format is then to allow the individual to make choices from refined options within the option last chosen. That is, if someone decides to switch from option A to option B between probability values of 0.1 and 0.2, the next stage of an iMPL would then prompt the subject to make more choices *within* this interval, to refine the values elicited. The probability intervals are 1 percentage point wide in the final stage of the iMPL, and the iteration stops at this threshold.¹²

As the subject iterates in the iMPL the choices become more and more alike, by design. Hence one would expect that greater cognitive effort would be needed to discriminate between them. At some point we expect the subject to express indifference.¹³ The iMPL uses the same incentive logic as the MPL. After the subjects have made all their choices, one row in each stage of the iMPL is selected at random by the experimenter.

A natural concern with the MPL is that it might encourage subjects to pick a response in the

¹² If the subject always chooses A or indicates indifference for any of the decision rows, there are no additional decisions required and the task is completed. Furthermore, the iterative format has some “smarts” built into it: when the values being elicited drop to some specified perceptive threshold (e.g., a 1-in-100 die throw), the iMPL collapses down to an endogenous number of final rows and the elicitation task stops iterating after those responses are entered.

¹³ In fact, one possible explanation of the observation that some subjects switch back and forth between choices in MPL is that they are indifferent. If so, explicitly including an indifference option, as we do here, may be a cleaner way to capture this behavior.

middle of the table, independent of true valuations. There could be a psychological bias towards the middle although that is not obvious *a priori*. The use of specific values at either end of the table could signal to the subject that the experimenter believes that these are reasonable upper and lower bounds. In some tasks, such as the risk elicitation task, the values are bounded by the laws of probability between 0 and 1, so this is less likely to be a factor compared to the pure psychological anchor of the middle row.

We test for framing effects in the field experiments by varying the cardinal scale of the MPL used, but only in the risk aversion tasks. In the laboratory experiments we vary the frame in both the risk aversion task and the discount rate task. Two asymmetric frames are developed: the *skewHI* treatment offers initial probabilities of (0.3, 0.5, 0.7, 0.8, 0.9 and 1), while *skewLO* offers initial probabilities of (0.1, 0.2, 0.3, 0.5, 0.7, and 1). This treatment yields 6 decision rows in Level 1 of the iMPL, as opposed to the 10 rows in the symmetric frame.¹⁴ As suggested by the treatment names, *skewLO* (*skewHI*) is intended to skew responses to be lower (higher) probabilities if subjects pick in the middle.

B. Individual Discount Rates

The experimental design for eliciting individual discount rates was introduced in Coller and Williams (1999) and extended by Harrison et al. (2002). The basic question used to elicit individual discount rates is simple: do you prefer 100 dollars today or $100+x$ dollars tomorrow, where x is some positive amount? If a risk-neutral subject prefers the 100 dollars today, then we can infer that the discount rate is higher than $x\%$ per day; otherwise, we can infer that it is $x\%$ per day or less.¹⁵ The format of our experiments extended this basic question in several ways.

¹⁴ The skewed frames interact with the implementation of the iMPL. In the symmetric frame, all intervals are 10 percentage points wide, so a second level is all that is needed to bring subject choices down to precise intervals of 1 percentage point. In the skewed frames, however, because the intervals vary in size, a third level is required to bring choices down to this level of precision, and the number of decision rows in Level 3 depends on the width of the interval in Level 1 at which the subject switches.

¹⁵ Andersen, Harrison, Lau and Rutström (2008a) show how one can extend the analysis of these data to make *joint* inferences about the concavity of the utility of income and discount rates. Given the evidence of risk aversion in these subjects, the effect is to lower the implied estimates of discount rates from those reported in Harrison, Lau and Williams (2002) significantly.

First, a number of such questions were posed to each individual, each question varying x by some amount. When x is zero we would obviously expect the individual to reject the option of waiting for no rate of return. As we increase x we would expect more individuals to take the future income option. For any given individual, the point at which they switch from choosing the current income option to taking the future income option provides a bound on their discount rate. That is, if an individual takes the current income option for all x from 0 to 10, then takes the future income option for all x from 11 up to 100, we can infer that their discount rate lies between 10% and 11% for this time interval. This inference assumes that the individual does not face perfect capital markets.¹⁶

Second, the experiments provided choices between two *future* income options rather than one “instant income” option and one future income option. For example, one could offer 100 dollars in one month from now and $100+x$ dollars in 7 months from now, interpreting the revealed discount rate as applying to a time horizon of 6 months. This avoids the potential problem of the subject facing extra transactions costs, including the possibility of default by the experimenter, with the future income option. If the delayed option were to involve greater transactions costs, then the revealed discount rate would include these subjective transactions costs. By having both options entail future income we hold these transactions costs constant.

Subjects in the field experiments were presented with six individual discount rate tasks in the same order, corresponding to six different time horizons: 1 month, 4 months, 6 months, 12 months, 18 months, and 24 months. Subjects in the laboratory experiments were presented with three tasks in random order, and the time horizons were 1 month, 4 months, and 6 months. The principal was 3000 kroner in all tasks, and payoffs to any one subject in the field experiments could range from 3000 kroner up to 7697 kroner, and this range converts to 458 and 1175 dollars.

The payoff table for the 6-month horizon treatment is illustrated in Table 2. Subjects were

¹⁶ If the individual did face perfect capital markets, then these decisions would simply reveal the market interest rate. This assumption is discussed in depth by Coller and Williams (1999) and Harrison et al. (2002). We collected detailed information from each subject to verify that their subjective borrowing and lending rates were different. When we allow for the formal possibility that the responses we obtain to our choice questions are censored by these field trading opportunities, we find similar qualitative results.

asked to choose between payment Options A and B for each of the 10 payoff alternatives. An option to express indifference was also included, although not illustrated in Table 2. Option A was 3000 kroner in all tasks, payable in 1 month. Option B paid $3000 + x$ kroner in 7 months, where x ranged from annual rates of return of 5% to 50% on the principal of 3000 kroner, compounded quarterly to be consistent with general Danish banking practices on overdraft accounts. The payoff tables provided the annual and annual effective interest rates for each option and the experimental instructions defined these two terms by way of example. We allow more refined elicitation of individual discount rates, and the annual interest rate intervals are 0.5% wide in the final stage of the iterative multiple price list.

3. Design of the Field and Laboratory Experiments

A. Field Experiments

The sample for the field experiments was designed to generate a representative sample of the adult Danish population. There were six steps in the construction of the sample, essentially following those employed in Harrison et al. (2002):

- First, a random sample of 25000 Danes was drawn from the Danish Civil Registration Office in January 2003. Only Danes born between 1927 and 1983 were included, thereby restricting the age range of the target population to between 19 and 75. For each person in this random sample we had access to their name, address, county, municipality, birth date, and sex. Due to the absence of names or addresses, 28 of these records were discarded.
- Second, we discarded 17 municipalities (including one county) from the population, due to them being located in extraordinarily remote locations. The population represented in these locations amounts to less than 2% of the Danish population, or 493 individuals in our sample of 25000 from the Civil Registry.
- Third, we assigned each county either 1 or 2 sessions, in rough proportionality to the population of the county. In total we assigned 20 sessions. Each session consisted of two sub-sessions at the same locale and date, one at 5pm and another at 8pm, and subjects were

allowed to choose which sub-session suited them best.

- Fourth, we divided 6 counties into two sub-groups because the distance between some municipalities in the county and the location of the session would be too large. A weighted random draw was made between the two sub-groups and the location selected, where the weights reflect the relative size of the population in September 2002.
- Fifth, we picked the first 30 or 60 randomly sorted records within each county, depending on the number of sessions allocated to that county. This provided a sub-sample of 600.
- Sixth, we mailed invitations to attend a session to the sub-sample of 600, offering each person a choice of times for the session. People were told that they would be paid 500 DKK to cover travel costs and that they would have a 10% chance of winning a significant amount. Response rates were low in some counties, so another 64 invitations were mailed out in these counties to newly drawn subjects. Everyone who gave a positive response was assigned to a session, and our recruited sample was 268, corresponding to a response rate of 40%.

Attendance at the experimental sessions was extraordinarily high, including 4 persons who did not respond to the letter of invitation but showed up unexpectedly and participated in the experiment. Four persons turned up for their session, but were not able to participate in the experiments.¹⁷ These experiments were conducted in June 2003, and a total of 253 subjects participated in the experiments.¹⁸

To minimize travel times for subjects, we reserved hotel meeting rooms in convenient locations across Denmark in which to conduct sessions.¹⁹ Participants met in groups of no more

¹⁷ The first person suffered from dementia and could not remember the instructions; the second person was a 76 year old woman who was not able to control the mouse and eventually gave up; the third person had just won a world championship in sailing and was too busy with media interviews to stay for two hours; and the fourth person was sent home because she arrived after the instructions had begun and we had already included one unexpected “walk-in” to fill their position.

¹⁸ Certain events might have plausibly triggered some of the no-shows: for example, 3 men did not turn up on June 11, 2003, but that was the night that the Danish national soccer team played a qualifying game for the European championships against Luxembourg that was not scheduled when we picked session dates.

¹⁹ It is possible to undertake experiments over the web with a large sample of subjects drawn from the population. Kapteyn and Teppa (2003) illustrate how one can elicit hypothetical responses to elicit time preferences using a panel of 2,000 Dutch households connected by home computer to surveys. Although not concerned with risk and time preferences directly, Hey (2002) illustrates how one can augment such electronic panel surveys with real experiments. Donkers and van Soest (1999) elicit hypothetical risk and time preferences from pre-existing panels of

than 10. To conduct computerized experiments in this field study, we found it was cost-effective to purchase laptop computers and transport them to the meeting sites. It was not necessary to network the computers for these experiments; the program ran independently on each computer, and results for each subject were saved onto the laptop that he or she used. Each subject is identified by a unique ID number in the data. For the randomization procedures, two bingo cages were used in each session, one containing 100 balls, and the other containing 3 to 11 balls, depending on the number of decision rows in the iMPL used in different treatments. We found two bingo cages to be the most transparent and convenient way to generate random outcomes in the experiments.

Subjects were first presented with the four separate risk aversion tasks, following a series of training tasks. The tasks were presented in the same order and differed in the prizes offered. After subjects had completed the four tasks, several random outcomes were generated to determine payments to subjects. For all subjects, one of the four tasks were chosen, then one of the decision rows in that task was chosen. For those subjects whose decision at that row led to the second stage of the iMPL, another random draw was required to choose a decision row at that stage, and yet another random draw was required should that decision have led a subject to the third stage in the iMPL. To maintain anonymity we performed the random draws in public without announcing to which subjects it would apply. In the case where a subject indicated indifference for the chosen decision row, another random draw determined whether the subject received the payments from Lottery A or Lottery B. At this point all subjects knew whether they were playing Lottery A or Lottery B, and another random draw determined if the subjects were to receive the high payment or the low payment. Finally, a 10-sided die with numbers from 0 up to 9 was rolled for each subject individually. Any subject who received a roll of “0” received actual payment according to the final outcome. All payments were made at the end of the experiment.

A significant amount of time was training subjects on the iMPL and the randomization procedures of the experiment. Subjects were presented with a trainer task in which the prizes in the iMPL were in the form of candies. The ten-sided die was rolled for each subject, and candies were

Dutch households being surveyed for other reasons.

given to each subject who received a roll of “0.”

In the second part of the experiment, subjects were presented with six individual discount rate tasks in the same order, corresponding to the six different time horizons. Future payments were guaranteed by the Danish Ministry of Economic and Business Affairs and made by automatic transfer from the Ministry’s bank account to the subject’s bank account. We applied the same randomization procedures as described above, and any subject who received a roll of “0” received actual payment according to the final outcome.

B. Laboratory Experiments

The laboratory experiments were conducted in October 2003. We recruited 100 subjects from the University of Copenhagen and the Copenhagen Business School. All subjects were recruited using the *ExLab* software.²⁰ The sessions were announced in 7 different lectures. At each lecture an announcement of the experiment was read aloud, and subjects were asked to enroll for the experiment by accessing *ExLab* through the Danish web page for this project. Of the 100 subjects recruited, 90 showed up for the experiment evenly spread across 9 sessions. Although several non-students participated, 74 out of the 90 subjects were students. Ages varied from 18 to 32 years, averaging 22.7 years, and only 27% were female.

We examined the performance of three MPL institutions (MPL, sMPL and iMPL) for each of the three framing conditions (“skew low”, “symmetric” and “skew high”) and the two types of valuation tasks (risk aversion and individual discount rates).²¹ We restrict the horizons in the discount rate task to three: 1, 4, and 6 months. The Switching MPL (sMPL) varies the standard MPL by asking the subject to choose which row he wants to switch at, assuming underlying monotonicity of the underlying preferences to fill out the remaining choices for the subject.²² The sMPL is

²⁰ This recruitment software is available for academic use at <http://exlab.bus.ucf.edu>. In addition, all instructions for our experiments are provided for public review at the *ExLab* Digital Library at the same location.

²¹ In the discount rate task, the *skewHI* treatment offers initial annual interest rates of 15%, 25%, 35%, 40%, 45%, and 50%, while the *skewLO* treatment offers annual interest rates of 5%, 10%, 15%, 25%, 35%, and 50%. The symmetric treatment offers 10 rows with annual interest rates between 5% and 50%.

²² We believe that the first implementation of the enforced-single-switching feature of the sMPL was by Gonzales and Wu (1999).

implemented because the iMPL changes the decision from the MPL in two ways: forcing a single switch point, and refining the choice. By comparing the MPL and sMPL we can see the pure effect of the first change, and by comparing the sMPL and the iMPL we can see the pure effect of the second change. Thus we have a $3 \times 3 \times 2$ design. The two first treatments are implemented between-subjects, so that any one subject experiences only one of the three MPL institutions and one of the three frames. The last treatment is implemented within-subjects, such that each subject faces 4 risk aversion tasks with varying stakes and 3 discount rate tasks with varying horizons. The monetary incentives in the valuation tasks are the same as those applied in the field experiments.

In addition, we examined two treatments that are applied equally to all subjects in each session. One is a randomization of their initial endowment. Each subject received a guaranteed 250 DKK to participate, and we randomly assigned them an extra amount between 10 and 100 DKK, chosen from a discrete uniform distribution in increments of 10 DKK. The second treatment is a randomization of the order of presentation of each of the four risk aversion tasks and each of the three individual discount rate (IDR) tasks. The rationale for these treatments and estimates of their effects are presented in Andersen et al. (2006).

4. Results

We examine the effect of our treatments on average measures of risk aversion and individual discount rates elicited in the field and laboratory experiments. Since we use the same valuation tasks and collect the same information on socio-demographic variables, we can compare correlations between treatments and demographic characteristics across the two samples. Detailed descriptions of the full range of treatments in the field and laboratory experiments are reported in Harrison et al. (2005c) and Andersen et al. (2006), respectively.

We examine the “overlap” between the two samples, which are the skewness frames in the risk aversion experiments and the three shorter time horizons in the discount rate experiments. In each case we retain controls for the additional treatments employed in the laboratory experiments (e.g., randomization of task order and use of random initial endowments). Similarly, there are some

demographic characteristics we simply did not observe in the laboratory but did observe in the field. For example, none of our laboratory subjects were retired, and none were older than 32. Again, we focus on the overlapping characteristics and control for others when examining the field data.

Our hypotheses concern three aspects of the elicited value. First, are the average elicited values the same in the field and the laboratory? Second, are the estimates of demographic effects the same in the field and the laboratory? In other words, do we estimate the same effect of observable characteristics such as sex or age on the elicited value? Finally, are the estimates of treatments effects the same in the field and the laboratory? Do we see the same effect of skewness frames on elicited risk attitudes, or the same effect of time horizon on elicited discount rates?

A. Measures of Risk Attitudes

Figure 1 shows the observed distribution of risk attitudes in our field and laboratory experiments, using the raw mid-point of final iterations of the elicited CRRA intervals. For comparability, the distributions reflect only the symmetric menu treatment. Using CRRA as the characterization of risk attitudes, a value of 0 denotes risk neutrality, negative values indicate risk-loving, and positive values indicate risk aversion. We see no significant difference in the average degree of risk aversion between the field and laboratory: the mean CRRA coefficient in the field sample is 0.63 with a 95% confidence interval between -0.49 and 1.87, while the mean coefficient is 0.79 in the laboratory sample with a 95% confidence interval between -0.02 and 1.85.²³

Although it is comforting to see that average risk aversion estimates are comparable in the

²³ The benchmark payoffs in Holt and Laury (2002) are 2 and 1.60 dollars in Option A and 3.85 and 0.10 dollars in Option B. They consider three treatments where the payoffs are scaled up by a factor 20, 50 and 90, with a CRRA of 0.54, 0.78 and 1.03, respectively, for the non-hypothetical responses. CRRA for their benchmark payoff is only 0.28. Harrison et al. (2005b) use a similar design and scale up the payoffs by a factor 10, and they find a CRRA coefficient of 0.57. Our payoffs are approximately equal to 150 times the benchmark payoffs in Holt and Laury (2002). Nevertheless, since we paid each subject with only a 10% probability, this could be argued to lower our payoff scale to a factor 15. We conducted a small experiment to examine directly the hypothesis that paying subjects with a 1-in-10 probability generates the same responses as paying them for certain. The control condition was a Holt and Laury (2002) design with lottery payoffs that were 20 and 16 dollars in Option A and 38.50 and 1 dollars in Option B. In the treatment condition each subject was given a 10% chance of being paid, and we rolled a 10-sided die for each of them individually to determine if they were to be paid. We recruited 77 subjects from the University of Central Florida student population: 51 in the control treatment and 26 in the 1-in-10 treatment. Using interval regression methods to estimate the effect of the treatment on elicited risk attitudes, we find that it is associated with a CRRA that is 0.11 lower than the control, but with a standard error of 0.13. Hence the p -value on the hypothesis test that there is no effect on risk attitudes from paying subjects with a 10% chance is 0.39, and we cannot reject the null hypothesis.

laboratory and the field, that does not address the variation with observable demographic characteristics and/or treatments. In order to assess the importance of demographics on risk attitudes, we apply regression models that condition on observable characteristics of the subjects. Table 3 provides the definitions of the explanatory variables and summary statistics of the two samples. The laboratory sample is much younger, which is no surprise. Similarly, we see large differences in characteristics such as whether the subjects have any children, whether they own their apartment or house, whether they are students, the completed level of education, income levels, or reside in greater Copenhagen. We also see more men and fewer women in the laboratory sample compared to the field sample: 51% of the field sample were female, compared to only 27% of the laboratory sample.

We apply interval regression models of elicited CRRA and IDR intervals that allow for the responses of the same subject to be correlated due to unobserved individual effects. The use of clustering to allow for “panel effects” from unobserved individual effects is common in the statistical survey literature. Clustering commonly arises in national field surveys from the fact that physically proximate households are often sampled to save time and money, but it can also arise from more homely sampling procedures. For example, Williams (2000, p.645) notes that it could arise from dental studies that “collect data on each tooth surface for each of several teeth from a set of patients” or “repeated measurements or recurrent events observed on the same person.” The procedures for allowing for clustering allow heteroskedasticity between and within clusters, as well as autocorrelation within clusters. They are closely related to the “generalized estimating equations” approach to panel estimation in epidemiology (see Liang and Zeger 1986), and generalize the “robust standard errors” approach popular in econometrics (see Rogers 1993).

Results from the Field

Table 4 displays the results from maximum likelihood estimation of a sample selection model of elicited risk attitudes, as well as a comparable interval regression model that does not allow for sample selection. Both sets of estimates allow for the possibility that observations on choices

made by the same subject are not independent. We find evidence of significant sample selection into the field experiments. The ancillary parameter ρ measures the estimated correlation between the residuals of the sample selection equation and the main CRRA equation. It is equal to 0.48, has a standard error of 0.24, and has a 95% confidence interval with values of +0.01 and +0.94. If this correlation had been zero, there would have been no evidence of sample selection bias on the main estimates of CRRA. The coefficients in the sample selection equation are jointly significant, as are many of the individual coefficients.

Turning to the variation in risk aversion across our field sample, we consider the marginal effect of individual demographics, holding constant the average value of other demographics. We find an effect on risk attitudes from age and education. Younger individuals, under the age of 30, tend to be more risk averse than those aged between 30 and 39, although the effect is not statistically significant. After the age of 40 subjects become significantly less risk averse than those aged between 30 and 39, the omitted age group.²⁴ We can convincingly reject the hypothesis that all age effects are the same (p -value is 0.001), as well as the null hypothesis that they are jointly zero (p -value is 0.001). Skilled workers, who are those subjects with some completed post-secondary education, have significantly higher aversion to risk than unskilled workers, and this effect on risk aversion is amplified if the subject has completed substantial higher education.

We also find that students generally exhibit significantly higher risk aversion than non-students, a result that is notable since we control for age effects and other characteristics that are correlated with being a student. Since the subjects were 19 years of age or older, they were all post-secondary students like most of the subjects attending our laboratory experiments. We thus observe that subjects with a completed post-secondary education, or a higher education, have higher aversion to risk than unskilled workers, and post-secondary students have higher risk aversion than non-students. Beginning or having completed post-secondary education is associated with higher aversion to risk than otherwise, which suggests that risk preferences may be in the same range if a

²⁴ We are not aware of other field experiments with monetary incentives that consider the correlation between individual risk attitudes and age. Holt and Laury (2002) report no significant age effect on risk attitudes using a sample of university students and faculty.

subject's "state of nature" changes from being unskilled and a student to having completed a post-secondary education or higher.²⁵

There is no significant effect from sex on risk attitudes. The absence of an effect of sex is noteworthy since it has been intensively studied using related experimental and survey methods and has even been the focus of theorizing about the role of evolution in forming preferences.²⁶

The variables *skewLO* and *skewHI* control for the frame used. The *skewHI* treatment is statistically significant, but there is no significant effect from the *skewLO* treatment.²⁷ The *skewLO* treatment yields an average estimated CRRA of 0.39, and the *skewHI* treatment yields an average estimated CRRA of 0.93, each in a direction expected *a priori*. Both estimates are consistent with the conclusion that subjects in the field experiments are risk averse.

We next look at interaction effects between treatment effects and two socio-demographic characteristics: age and sex. We pick these two since there is considerable variation in sex within both samples, and there is not much variation in age within the laboratory sample. To examine these interaction effects we re-estimate the regression models reported in Table 5 for the field sample and then test for the absence of interaction effects formally.²⁸ We reject the hypothesis that both males and females provided the same CRRA value independently of frame (p -value of 0.001 and 0.022,

²⁵ Andersen et al. (2008b) find some variation in risk attitudes over time, but there is no general tendency for risk attitudes to increase or decrease over a 17-month span. The results also suggest that risk preferences are state contingent with respect to personal finances. Using similar valuation tasks but smaller monetary incentives, Harrison et al. (2005a) find that risk attitudes for the same individuals are stable over the six month period used in their laboratory experiments. They do not consider changes in states of nature since the time period is relatively short.

²⁶ Levin et al. (1988) and Powell and Ansic (1997) illustrate the experimental studies undertaken in a settings in which the task was not abstract but there were no real earnings by subjects. Harbaugh et al. (2002) and Holt and Laury (2002) conduct abstract experiments with real rewards, and find no significant sex effects on elicited risk aversion when stakes are non-trivial. Schubert et al. (1999) conduct abstract and non-abstract experiments with real rewards and conclude that women do appear to be more risk averse than men in abstract tasks in the gain frame, but that this effect disappears with context. Unfortunately, they employed the Becker-DeGroot-Marschak procedure for eliciting certainty-equivalents, which is known to have poor incentive properties for experimental subjects. Jianakoplos and Bernasek (1998) examine data from the *U.S. Survey of Consumer Finances* and conclude that single women are more risk averse in their financial choices than single men. Eckel and Grossman (2008b) review these studies and several unpublished studies. Rubin and Paul (1979) and Robson (1996) offer evolutionary models of possible sex differences in risk aversion. Even if there is no evidence for an effect of sex on risk aversion, it is possible that observers may *predict* differences in risk attitudes based on sex (Eckel and Grossman 2008a).

²⁷ Harrison et al. (2005c) show that there is a significant effect on elicited risk attitudes from *both* skewness treatments on the *initial* stage of the iMPL, but that the iterations of the iMPL make that effect disappear for the *skewLO* treatment. Our analysis focuses on the final stage of the iMPL procedure. Thus, one should be concerned about the possible effects of such framing on eliciting risk attitudes if using the original MPL procedure.

²⁸ The detailed statistical results are reported in the output files available at the *ExLab* Digital Library at <http://exlab.bus.ucf.edu> and are not of interest beyond the tests of interaction effects reported in the text. Each test is a χ^2 test that the appropriate coefficient estimates are identical.

respectively). Although borderline in terms of conventional significance levels, we would also reject the hypothesis that males and females provided the same CRRA value within the symmetric frame at a p -value of 0.094. Thus there is evidence of interaction effects between sex and the treatments from the field data.

Turning to age, we find more evidence of interaction effects. Within the *skewLO* frame, we reject the hypothesis that age does not matter with a p -value of 0.027; within the symmetric frame the same test generates a p -value of less than 0.001. Particularly telling is the finding that these effects are driven by the older subjects in the field. We reject the hypothesis that middle-aged subjects, between 40 and 50, have the same CRRA when faced with different frames at a p -value of 0.003. The same test using subjects aged over 50 is rejected at a p -value of less than 0.001.

Results from the Laboratory

Table 5 reports results from an interval regression model of elicited CRRA values in our laboratory experiments, controlling for MPL institution, framing condition, task order, prize set, experimenter and individual demographics.²⁹ Unobserved individual effects are again modeled using the clustering specification.

In general, we see a substantially smaller effect of heterogeneity in our laboratory sample compared to the field sample. The only significant demographic variable is “single” status. The elicited CRRA value for singles is on average 0.26 lower than subjects with a partner, and this effect has a p -value of 0.02. We do not find a significant effect from age or education on elicited CRRA values in the laboratory sample, no doubt due to the homogeneity of the sample: 98% of the subjects are less than 30 years of age, 82% are students, and 78% have not completed some post-secondary education and are classified as unskilled workers.

Turning to estimates of treatment effects, we see that the *skewLO* frame lowers average CRRA by about 0.23, and this effect is statistically significant with a p -value of 0.07. In the field this

²⁹ There are no observations on the two highest age groups and the retired group in our laboratory experiments, and three binary variables related to these demographics are therefore excluded in the model.

effect was to lower average CRRA by 0.15. and the p -value on that estimated effect was 0.11. We also find that the *skewHI* frame has no effect on average CRRA in the laboratory (the coefficient estimate is only 0.03 with a p -value of 0.80), compared to a significant increase of 0.28 in the field (with a p -value of less than 0.01). Thus subjects across laboratory and field samples respond differently to the framing treatments, but the effects are more or less consistent with the hypothesis that motivated the treatments.

Looking at interaction effects between treatment effects and sex in the laboratory sample, we can reject the hypothesis that females provided the same CRRA value independently of the frame (p -value of 0.068). However, we cannot reject the hypothesis that men provide the same CRRA value across the three frames (p -value of 0.56). We find that both women and men in the field sample provided different CRRA values across frames. Hence, we find different interaction effects between treatment effects and sex across the field and laboratory samples.

B. Measures of Individual Discount Rates

Figure 2 displays the elicited discount rates for our subjects in the field and laboratory experiments using the mid-point of the final interval selected. The discount rates are pooled across all horizons included in the laboratory experiments (1, 4, and 6 months), and these distributions only reflect the symmetric menu treatment in the laboratory.³⁰ The mean elicited discount rate in the field experiments is 25.0%, with a 95% confidence interval between 0.9% and 51.7%.³¹ The mean coefficient in the laboratory experiments is 27.9%, and a 95% confidence interval between 5.0% and 47.7%. Thus we see similar average measures of individual discount rates when we compare the laboratory and the field samples.

³⁰ We did not employ framing treatments in the field IDR experiments, so using the symmetric laboratory results is directly comparable to the field results.

³¹ These values are virtually identical to those found in the field by Harrison et al. (2005c): they estimated a mean of 23.1%, a median of 22.4%, and a standard deviation of 14.0%.

Results from the Field

Table 6 displays the results from the maximum likelihood estimation of a sample selection model of elicited IDR values and a comparable interval regression model that does not allow for sample selection. Both sets of estimates allow for the possibility that observations on choices made by the same subject are not independent. Contrary to the risk experiments, there is no evidence of significant sample selection into the field experiments. The ancillary parameter ρ is equal to -0.20, has a standard error of 0.37, and has a 95% confidence interval with values of -0.93 and +0.53. The coefficients in the sample selection equation are jointly significant, as are a few of the individual coefficients.

The statistical model includes the 1-month, 4-months, and 6-months time horizons to facilitate comparisons with the laboratory experiments. The default horizon is the 1-month task, so we observe that there is a small but statistically significant effect of horizon on elicited discount rates. The 4-month and 6-month discount rates are approximately 3 percentage points lower than the 1-month average. None of the demographic characteristics are significant at conventional levels in the model that corrects for sample selection. The only demographic characteristic to have a statistically significant effect in the model with no sample selection correction is living in the Copenhagen area, which increases average discount rates by 5.5 percentage points (p -value of 0.05). The related variable capturing if someone lives in a larger city of 20,000 inhabitants or more also increases average IDR by 4.1 percentage points (with a p -value of 0.08). Higher education and low income have a less significant effect, the former decreasing average IDR by about 5 percentage points and the latter increasing it by about the same amount. Each is significant at roughly the 10% level. Although heterogeneity is an issue for discount rates, it does not appear to be as strong as it is for risk attitudes.³²

³² Although not a focus of our analysis, it is interesting to note that there is a significant experimenter effect in the regression model that corrects for sample selection. It appears that experimenter Andersen was associated with elicited discount rates that were 3.65 percentage points lower than experimenter Lau, and this effect has a p -value of 0.07. This could be due to Lau being a little slower at conducting the experiments: over a session that generally required 2 hours, he typically required 15 to 20 minutes more. This could have encouraged some impatience in subjects that spilled over into their responses, although this is conjecture. The same differences in speed of executing the experiment were found in the lab. The average IDR is again lower with Andersen at the helm by 7.18 percentage points and with a p -value of 0.08.

Turning to estimates of interaction effects, we again find that individual characteristics affect the estimated responses to the treatments. The most interesting result is that significant horizon differences are observed only for women. We reject the hypothesis that females provided the same IDR independently of horizon (p -value of 0.013). Although borderline in terms of conventional significance levels, we would also reject the hypothesis that adults aged between 30 and 39 provide the same IDR across horizons at a p -value of 0.084. Thus there is evidence of interaction effects between sex and the treatments from the field data.

Results from the Laboratory

Table 7 displays comparable results from the tasks in the laboratory experiments. We find that the treatment effects are roughly double those obtained in the field, and they are statistically significant. Although no demographic effect is statistically significant at the 10% level, many have large coefficients. For example, those subjects who have completed higher education have discount rates that are 8.3 percentage points lower than otherwise, those subjects who own their house or apartment have discount rates that are lower by 7.3 percentage points, and the coefficient for those subjects living in Copenhagen is minus 8.4 percentage points. Finally, we reject the hypothesis that both women and men provided the same IDR independently of horizon (p -value of 0.078 and less than 0.001, respectively).

5. Conclusions

Our results provide evidence that there are good reasons to conduct field and laboratory experiments. We find no significant difference in the average degree of risk aversion and discount rates between the field and laboratory samples. However, we find evidence of preference heterogeneity based on our field responses, which implies that a change in the composition of the laboratory sample may result in responses that on average differ from those elicited in the field. The homogeneity of the university student population limits the ability of laboratory experiments to detect the preference heterogeneity that is present in the broader population. We also find

differences in treatment effects measured in the laboratory and the field that can, at least in part, be traced to interactions between treatment and demographic effects.

Our results apply to experiments in which we elicit individual risk attitudes and discount rates. One might expect such experiments to exhibit more variation across individual characteristics than many laboratory experiments in which preferences over risk and time play a less important role. For example, it is not obvious that our results would cause one to doubt the vast set of experimental results on the relative efficiency of double auction markets relative to posted offer markets.

References

- Andersen, S., Harrison, G. W., Lau, M. I., Rutström, E. E., 2006. Elicitation using multiple price list format. *Experimental Economics*, 9, 383-405.
- Andersen, S., Harrison, G. W., Lau, M. I., Rutström, E. E., 2008a. Eliciting risk and time preferences. *Econometrica*, 76, 583-618.
- Andersen, S., Harrison, G. W., Lau, M. I., Rutström, E. E., 2008b. Lost in state space: Are preferences stable?. *International Economic Review*, 49, 1091-1112
- Ballinger, T. P., Wilcox, N. T., 1997. Decisions, error and heterogeneity. *Economic Journal*, 107, 1090-1105.
- Blackburn, M., Harrison, G. W., Rutström, E. E., 1994. Statistical bias functions and informative hypothetical surveys. *American Journal of Agricultural Economics*, 76, 1084-1088.
- Camerer, C., 2003. *Behavioral game theory*. New Jersey: Princeton University Press.
- Coller, M., Williams, M. B., 1999. Eliciting individual discount rates. *Experimental Economics*, 2, 107-127.
- Cox, J. C., Sadiraj, V., 2006. Small- and large-stakes risk aversion: Implications of concavity calibration for decision theory. *Games & Economic Behavior*, 56, 45-60.
- Dehejia, R. H., Wahba, S., 1999. Causal effects in nonexperimental studies: Reevaluating the evaluation of training programs. *Journal of The American Statistical Association*, 94, 1053-62.
- Dehejia, R. H., Wahba, S., 2002. Propensity score matching for nonexperimental causal studies. *Review of Economics & Statistics*, 84, 151-61.
- Dohmen, T., Falk, A., Huffman, D., Sunde, U., Schupp, J., Wagner, G. G., 2005. Individual risk attitudes: New evidence from a large representative, experimentally-validated survey. Discussion Paper 1730, Institute for the Study of Labor, Bonn, Germany.

- Donkers, B., van Soest, A., 1999. Subjective measures of household preferences and financial decisions. *Journal of Economic Psychology*, 20, 613-642.
- Eckel, C. C., Grossman, P. J., 2008a. Forecasting risk attitudes: An experimental study using actual and forecast gambling choices. *Journal of Economic Behavior & Organization*, 68(1), 1-17.
- Eckel, C. C., Grossman, P. J., 2008b. Sex and risk: Experimental evidence. In C.R. Plott and V.L. Smith (eds.), *Handbook of Results in Experimental Economics*. Amsterdam: North Holland/Elsevier Press
- Eckel, C., Johnson, C., Montmarquette, C., 2005. Savings decisions of the working poor: Short- and long-term horizons. In J. Carpenter, G.W. Harrison and J.A. List (eds.), *Field Experiments in Economics*. Greenwich, CT: JAI Press, Research in Experimental Economics.
- Gonzalez, R., Wu, G., 1999. On the shape of the probability weighting function. *Cognitive Psychology*, 38, 129-166.
- Harbaugh, W. T., Krause, K., Vesterlund, L., 2002. Risk attitudes of children and adults: Choices over small and large probability gains and losses. *Experimental Economics*, 5, 53-84.
- Harrison, G. W., 2006 Experimental evidence on alternative environmental valuation methods. *Environmental and Resource Economics*, 34, 125-162.
- Harrison, G. W., Jensen, J., Lau, M. I., Rutherford, T. F., 2002a. Policy reform without tears. In A. Fossati and W. Weigard (eds.), *Policy Evaluation With Computable General Equilibrium Models*. New York: Routledge.
- Harrison, G. W., Johnson, E., McInnes, M. M., Rutström, E. E., 2005a. Temporal stability of estimates of risk aversion. *Applied Financial Economics Letters*, 1, 31-35.
- Harrison, G. W., Johnson, E., McInnes, M. M., Rutström, E. E., 2005b. Risk aversion and incentive effects: Comment. *American Economic Review*, 95, 897-901.
- Harrison, G. W., Lau, M. I., Rutström, E. E., 2007, Estimating risk attitudes in Denmark: A field experiment. *Scandinavian Journal of Economics*, 109, 341-368.
- Harrison, G. W., Lau, M. I., Rutström, E. E., Sullivan, M. B., 2005c. Eliciting risk and time preferences using field experiments: Some methodological issues. in J. Carpenter, G.W. Harrison and J.A. List (eds.), *Field Experiments in Economics*. Greenwich, CT: JAI Press, Research in Experimental Economics.
- Harrison, G. W., Lau, M. I., Williams, M. B., 2002b. Estimating individual discount rates for Denmark: A field experiment. *American Economic Review*, 92, 1606-1617.
- Harrison, G. W., List, J. A., 2004. Field experiments. *Journal of Economic Literature*, 42, 1013-1059.
- Harrison, G. W., Vinod, H. D., 1992. The sensitivity analysis of applied general equilibrium models: Completely randomized factorial sampling designs. *Review of Economics and Statistics*, 74, 357-362.
- Hey, J. D., 2002. Experimental economics and the theory of decision making under uncertainty. *Geneva Papers on Risk and Insurance Theory*, 27, 5-21.
- Holt, C. A., Laury, S. K., 2002. Risk aversion and incentive effects. *American Economic Review*, 92,

1644-1655.

- Jianakoplos, N. A., Bernasek, A., 1998. Are women more risk averse?. *Economic Inquiry*, 36, 620-630.
- Kagel, J. H., Levin, D., 2002. *Common Value Auctions and the Winner's Curse*. New Jersey: Princeton University Press.
- Kapteyn, A., Teppa, F., 2003. Hypothetical intertemporal consumption choices. *Economic Journal*, 113, C140-C151.
- Lalonde, R. J., 1986. Evaluating the econometric evaluations of training programs with experimental data. *American Economic Review*, 76, 604-620.
- Lau, M. I., 2000. Assessing tax reforms when human capital is endogenous. In G. W. Harrison, S. E. H. Jensen, L. H. Pedersen and T. F. Rutherford (eds.), *Using Dynamic General Equilibrium Models for Policy Analysis*. Amsterdam: North Holland, Contributions to Economic Analysis 248.
- Levin, I. P., Snyder, M. A., Chapman, D. P., 1988. The interaction of experiential and situational factors and gender in a simulated risky decision-making task. *Journal of Psychology*, 122, 173-181.
- Liang, K-Y., Zeger, S.L., 1986. Longitudinal data analysis using generalized linear models. *Biometrika*, 73, 13-22.
- Powell, M., Ansic, D., 1997. Gender differences in risk behavior in financial decision-making: An experimental analysis. *Journal of Economic Psychology*, 18, 605-628.
- Rabin, M., 2000. Risk aversion and expected utility theory: A calibration theorem. *Econometrica*, 68, 1281-1292.
- Robson, A. J., 1996. The evolution of attitudes to risk: Lottery tickets and relative wealth. *Games and Economic Behavior*, 14, 190-207.
- Rogers, W. H., 1993. Regression standard errors in clustered samples. *Stata Technical Bulletin*, 13, 19-23.
- Rosenbaum, P., Rubin, D., 1983. The central role of the propensity score in observational studies for causal effects. *Biometrika*, 70, 41-55.
- Rosenbaum, P., Rubin, D., 1984. Reducing bias in observational studies using multivariate matched sampling methods that incorporate the propensity score. *Journal of the American Statistical Association*, 79, 39-68.
- Rubin, P. H., Paul, C. W., 1979. An evolutionary model of taste for risk. *Economic Inquiry*, 17, 585-595.
- Saha, A., 1993. Expo-power utility: A flexible form for absolute and relative risk aversion. *American Journal of Agricultural Economics*, 75, 905-913.
- Schubert, R., Brown, M., Gysler, M., Brachinger, H. W., 1999. Financial decision-making: Are women really more risk averse?. *American Economic Review (Papers & Proceedings)*, 89, 381-385.

- Smith, J., Todd, P., 2005. Does matching address LaLonde's critique of nonexperimental estimates?. *Journal of Econometrics*, 125, 305-353.
- Starmer, C., 2000. Developments in non-expected utility theory: The hunt for a descriptive theory of choice under risk. *Journal of Economic Literature*, XXXVIII, 332-382.
- United States Office of Management and Budget, 2006. Standards and guidelines for statistical surveys. (Washington, DC; Office of Management and Budget).
- Williams, R. L., 2000. A Note on robust variance estimation for cluster-correlated data. *Biometrics*, 56, 645-646.

Table 1: Payoff Matrix in the Holt and Laury Risk Aversion Experiments

Default payoff matrix for scale 1

Lottery A				Lottery B				EV ^A	EV ^B	Difference	Open CRRA Interval if Subject Switches to Lottery B
p(\$2)		p(\$1.60)		p(\$3.85)		p(\$0.10)					
0.1	\$2	0.9	\$1.60	0.1	\$3.85	0.9	\$0.10	\$1.64	\$0.48	\$1.17	$-\infty, -1.71$
0.2	\$2	0.8	\$1.60	0.2	\$3.85	0.8	\$0.10	\$1.68	\$0.85	\$0.83	-1.71, -0.95
0.3	\$2	0.7	\$1.60	0.3	\$3.85	0.7	\$0.10	\$1.72	\$1.23	\$0.49	-0.95, -0.49
0.4	\$2	0.6	\$1.60	0.4	\$3.85	0.6	\$0.10	\$1.76	\$1.60	\$0.16	-0.49, -0.15
0.5	\$2	0.5	\$1.60	0.5	\$3.85	0.5	\$0.10	\$1.80	\$1.98	-\$0.17	-0.15, 0.14
0.6	\$2	0.4	\$1.60	0.6	\$3.85	0.4	\$0.10	\$1.84	\$2.35	-\$0.51	0.14, 0.41
0.7	\$2	0.3	\$1.60	0.7	\$3.85	0.3	\$0.10	\$1.88	\$2.73	-\$0.84	0.41, 0.68
0.8	\$2	0.2	\$1.60	0.8	\$3.85	0.2	\$0.10	\$1.92	\$3.10	-\$1.18	0.68, 0.97
0.9	\$2	0.1	\$1.60	0.9	\$3.85	0.1	\$0.10	\$1.96	\$3.48	-\$1.52	0.97, 1.37
1	\$2	0	\$1.60	1	\$3.85	0	\$0.10	\$2.00	\$3.85	-\$1.85	1.37, ∞

Note: The last four columns in this table, showing the expected values of the lotteries and the implied CRRA intervals, were not shown to subjects.

Table 2: Payoff Table for 6 Month Time Horizon in the Discount Rate Experiments

Payoff Alternative	Payment Option A (pays amount below in 1 month)	Payment Option B (pays amount below in 7 months)	Annual Interest Rate (AR, in percent)	Annual Effective Interest Rate (AER, in percent)	Preferred Payment Option (Circle A or B)
1	3,000 DKK	3,075 DKK	5	5.09	A B
2	3,000 DKK	3,152 DKK	10	10.38	A B
3	3,000 DKK	3,229 DKK	15	15.87	A B
4	3,000 DKK	3,308 DKK	20	21.55	A B
5	3,000 DKK	3,387 DKK	25	27.44	A B
6	3,000 DKK	3,467 DKK	30	33.55	A B
7	3,000 DKK	3,548 DKK	35	39.87	A B
8	3,000 DKK	3,630 DKK	40	46.41	A B
9	3,000 DKK	3,713 DKK	45	53.18	A B
10	3,000 DKK	3,797 DKK	50	60.18	A B

Table 3: List of Variables and Descriptive Statistics

Variable Definition		Field Sample Mean	Laboratory Sample Mean
female	Female	0.51	0.27
young	Aged less than 30	0.17	0.98
middle	Aged between 40 and 50	0.28	0.00
old	Aged over 50	0.38	0.00
single	Lives alone	0.20	0.48
kids	Has children	0.28	0.02
owner	Owens own home or apartment	0.69	0.19
retired	Retired	0.16	0.00
student	Student	0.09	0.82
skilled	Some post-secondary education	0.38	0.11
longedu	Substantial higher education	0.36	0.14
IncLow	Lower level income	0.34	0.74
IncHigh	Higher level income	0.34	0.16
copen	Lives in greater Copenhagen area	0.27	0.87
city	Lives in larger city of 20,000 or more	0.39	0.09
experimenter	Experimenter Andersen (default is Lau)	0.49	0.66
County_15	Koebenhavns Amt	0.10	--
County_20	Frederiksborg Amt	0.09	--
County_25	Roskilde Amt	0.05	--
County_30	Vestsjaellands Amt	0.09	--
County_42	Fyns Amt	0.09	--
County_50	Soenderjyllands Amt	0.05	--
County_55	Ribe Amt	0.06	--
County_60	Vejle Amt	0.06	--
County_65	Ringkoebing Amt	0.11	--
County_70	Aarhus Amt	0.09	--
County_80	Nordjyllands Amt	0.10	--
wave2	Second wave of invitations	0.03	--
wave3	Third wave of invitations	0.02	--

Legend: Most variables have self-evident definitions. The omitted age group is 30-39. Variable “skilled” indicates if the subject has completed vocational education and training or “short-cycle” higher education, and variable “longedu” indicates the completion of “medium-cycle” higher education or “long-cycle” higher education. These terms for the cycle of education are commonly used by Danes (most short-cycle higher education program last for less than 2 years; medium-cycle higher education lasts 3 to 4 years, and includes training for occupations such as a journalist, primary and lower secondary school teacher, nursery and kindergarten teacher, and ordinary nurse; long-cycle higher education typically lasts 5 years and is offered at Denmark’s five ordinary universities, at the business schools and various other institutions such as the Technical University of Denmark, the schools of the Royal Danish Academy of Fine Arts, the Academies of Music, the Schools of Architecture and the Royal Danish School of Pharmacy). Lower incomes are defined in variable “IncLow” by a household income in 2002 below 300,000 kroner. Higher incomes are defined in variable “IncHigh” by a household income of 500,000 kroner or more.

Table 4: Statistical Model of Risk Aversion Responses in Field Experiments

Interval regression estimates of CRRA corrected for sample selection bias. N=846 and 242 subjects.

Variable	Variable Description	Sample Selection Correction			No Correction		
		Estimate	Standard Error	<i>p</i> -value	Estimate	Standard Error	<i>p</i> -value
<i>A. CRRA Equation</i>							
Constant		-0.25	0.23	0.27	0.22	0.18	0.24
skewLO	Frame to skew RA down	-0.15	0.09	0.11	-0.03	0.12	0.83
skewHI	Frame to skew RA up	0.28	0.08	0.00	0.25	0.09	0.01
PrizeSet2	Second prize set	0.29	0.06	0.00	0.29	0.06	0.00
PrizeSet3	Third prize set	0.21	0.05	0.00	0.20	0.05	0.00
PrizeSet4	Fourth prize set	0.17	0.04	0.00	0.19	0.05	0.00
experimenter	Experimenter effect	-0.09	0.08	0.22	-0.04	0.09	0.63
female	Female	-0.02	0.07	0.81	-0.05	0.09	0.54
young	Aged less than 30	0.24	0.16	0.14	0.15	0.18	0.40
middle	Aged between 40 and 50	-0.28	0.11	0.01	-0.30	0.13	0.03
old	Aged over 50	-0.21	0.12	0.09	-0.15	0.15	0.31
single	Lives alone	0.12	0.10	0.25	0.08	0.13	0.52
kids	Has children	0.03	0.10	0.73	0.05	0.12	0.68
owner	Owens home or apartment	0.20	0.09	0.02	0.07	0.11	0.51
retired	Retired	0.06	0.11	0.58	-0.08	0.12	0.50
student	Student	0.30	0.13	0.03	0.32	0.16	0.05
skilled	Some post-secondary education	0.28	0.08	0.00	0.29	0.10	0.00
longedu	Substantial higher education	0.33	0.10	0.00	0.37	0.12	0.00
IncLow	Lower level income	-0.01	0.09	0.90	0.01	0.12	0.95
IncHigh	Higher level income	0.02	0.09	0.79	0.02	0.12	0.85
copen	Lives in Copenhagen area	0.16	0.10	0.12	0.16	0.12	0.19
city	Lives in larger city of 20,000 or more	0.06	0.09	0.49	0.14	0.10	0.17
<i>B. Sample Selection Equation</i>							
Constant	Constant	0.74	0.15	0.00			
female	Female	-0.13	0.08	0.14			
young	Aged less than 30	0.20	0.13	0.13			
middle	Aged between 40 and 50	0.19	0.12	0.12			
old	Aged over 50	0.03	0.12	0.80			
County_15	Koebenhavns Amt	-0.23	0.17	0.19			
County_20	Frederiksborg Amt	-0.32	0.18	0.08			
County_25	Roskilde Amt	-0.39	0.22	0.07			
County_30	Vestsjaellands Amt	-0.53	0.18	0.00			
County_42	Fyns Amt	-0.30	0.18	0.09			
County_50	Soenderjyllands Amt	-0.35	0.21	0.10			
County_55	Ribe Amt	-0.45	0.22	0.04			
County_60	Vejle Amt	0.02	0.19	0.90			
County_65	Ringkoebing Amt	-0.05	0.18	0.80			
County_70	Aarhus Amt	-0.33	0.18	0.07			
County_80	Nordjyllands Amt	-0.40	0.19	0.03			
wave2	Second wave of invitations	-0.51	0.22	0.02			
wave3	Third wave of invitations	-0.33	0.35	0.36			
ρ	Error correlation	0.48	0.24	0.05			

Table 5: Statistical Model of Risk Aversion Responses in Laboratory Experiments

Interval regression estimates of CRRA.
N=254, based on 90 subjects.

Variable	Variable Description	Estimate	Standard Error	p-value
Constant		0.78	0.35	0.03
smpl	sMPL format	0.02	0.10	0.84
impl	iMPL format	0.38	0.12	0.00
skewLO	Frame to skew RA down	-0.23	0.12	0.07
skewHI	Frame to skew RA up	-0.03	0.10	0.80
Task2	Second task	0.05	0.05	0.28
Task3	Third task	0.03	0.05	0.51
Task4	Fourth task	0.14	0.05	0.01
PrizeSet2	Second prize set	0.08	0.06	0.14
PrizeSet3	Third prize set	0.06	0.04	0.17
PrizeSet4	Fourth prize set	0.03	0.04	0.36
experimenter	Experimenter effect	-0.10	0.10	0.33
endowment	Initial endowment	-0.00	0.00	0.26
female	Female	0.10	0.10	0.28
single	Lives alone	-0.26	0.11	0.02
owner	Has children	0.11	0.15	0.49
student	Owens home or apartment	-0.06	0.15	0.72
skilled	Some post-secondary education	-0.03	0.10	0.76
longedu	Substantial higher education	-0.06	0.14	0.66
IncLow	Lower level income	0.08	0.19	0.66
IncHigh	Higher level income	-0.06	0.21	0.79
copen	Lives in Copenhagen area	0.10	0.21	0.65
city	Lives in larger city of 20,000 or more	0.01	0.28	0.96

Table 6: Statistical Model of IDR Responses in Field Experiments

Interval regression estimates of IDR corrected for sample selection bias.
N=652, based on 236 subjects.

Variable	Variable Description	Sample Selection Correction			No Correction		
		Estimate	Standard Error	<i>p</i> -value	Estimate	Standard Error	<i>p</i> -value
<i>A. IDR Equation</i>							
Constant		26.28	6.70	0.00	28.82	5.45	0.00
horizon4	4 months horizon	-2.08	1.00	0.04	-3.28	1.04	0.00
horizon6	6 months horizon	-2.79	1.04	0.01	-3.69	1.14	0.00
experimenter	Experimenter effect	-3.65	2.01	0.07	-2.65	2.15	0.22
female	Female	1.97	1.93	0.31	0.48	2.14	0.82
young	Aged less than 30	-5.52	3.79	0.14	-4.49	4.31	0.30
middle	Aged between 40 and 50	-0.16	3.02	0.96	1.64	3.40	0.63
old	Aged over 50	3.19	3.22	0.32	3.92	3.51	0.26
single	Lives alone	1.65	2.36	0.48	0.28	2.88	0.92
kids	Has children	2.28	2.67	0.39	3.74	2.94	0.20
owner	Own home or apartment	-0.38	2.47	0.88	-0.36	2.85	0.90
retired	Retired	-0.47	2.57	0.85	-3.62	2.77	0.19
student	Student	-3.03	3.38	0.37	-1.35	4.42	0.76
skilled	Some post-secondary education	1.88	2.08	0.37	-1.65	2.63	0.53
longedu	Substantial higher education	-3.26	2.23	0.14	-4.94	2.89	0.09
IncLow	Lower level income	2.88	2.58	0.26	4.79	2.90	0.10
IncHigh	Higher level income	0.65	2.37	0.78	-1.17	2.69	0.66
copen	Lives in Copenhagen area	2.47	2.70	0.36	5.47	2.77	0.05
city	Lives in larger city of 20,000 or more	3.01	2.14	0.16	4.14	2.38	0.08
<i>B. Sample Selection Equation</i>							
Constant		0.85	0.15	0.00			
female	Female	-0.09	0.08	0.23			
young	Aged less than 30	0.23	0.13	0.08			
middle	Aged between 40 and 50	0.25	0.12	0.04			
old	Aged over 50	0.11	0.11	0.34			
County_15	Koebenhavns Amt	-0.07	0.17	0.66			
County_20	Frederiksborg Amt	-0.29	0.17	0.09			
County_25	Roskilde Amt	-0.27	0.23	0.23			
County_30	Vestsjaellands Amt	-0.35	0.18	0.06			
County_42	Fyns Amt	-0.21	0.19	0.29			
County_50	Soenderjyllands Amt	-0.38	0.22	0.08			
County_55	Ribe Amt	-0.26	0.21	0.22			
County_60	Vejle Amt	0.13	0.24	0.59			
County_65	Ringkoebing Amt	-0.06	0.17	0.70			
County_70	Aarhus Amt	-0.27	0.17	0.12			
County_80	Nordjyllands Amt	-0.24	0.17	0.17			
wave2	Second wave of invitations	-0.51	0.23	0.03			
wave3	Third wave of invitations	-0.44	0.35	0.21			
ρ	Error correlation	-0.20	0.37	0.59			

Table 7: Statistical Model of IDR Responses in Laboratory Experiments

Interval regression estimates of IDR.
N=270, based on 90 subjects.

Variable	Variable Description	Estimate	Standard Error	p-value
Constant		56.25	17.38	0.00
horizon4	4 months horizon	-6.47	1.75	0.00
horizon6	6 months horizon	-8.84	1.81	0.00
smpl	sMPL format	-3.67	4.85	0.45
impl	iMPL format	3.15	4.02	0.43
Task2	Second task	3.60	1.51	0.02
Task3	Third task	2.14	1.43	0.13
skewLO	Frame to skew IDR down	3.78	4.72	0.42
skewHI	Frame to skew IDR up	4.19	4.11	0.31
experimenter	Experimenter effect	-7.18	4.09	0.08
endowment	Initial endowment	-0.02	0.06	0.80
female	Female	2.68	3.78	0.48
single	Lives alone	-5.93	4.17	0.15
owner	Own home or apartment	-7.26	5.95	0.22
student	Student	-6.20	4.54	0.17
skilled	Some post-secondary education	-6.77	5.34	0.20
longedu	Substantial higher education	-8.32	5.36	0.12
IncLow	Lower level income	-4.39	5.79	0.45
IncHigh	Higher level income	-5.26	6.09	0.39
copen	Lives in Copenhagen area	-8.38	13.71	0.54
city	Lives in larger city of 20,000 or more	0.76	15.02	0.96

Figure 1: Distribution of CRRA in Denmark
Mid-Point of Raw Response Interval with Symmetric Menu

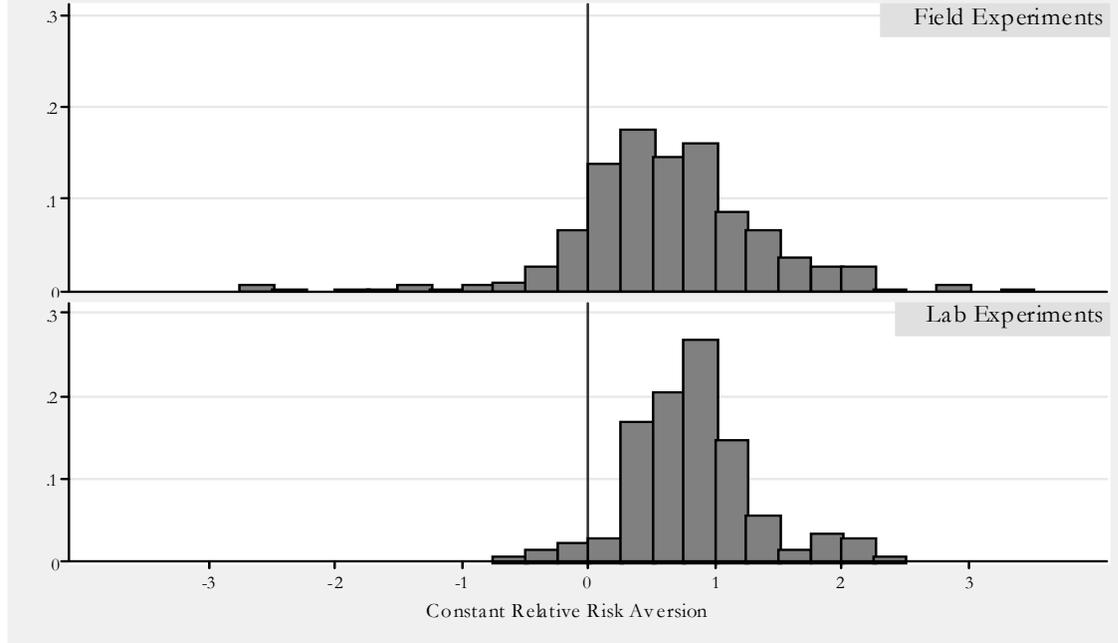


Figure 2: Distribution of IDR in Denmark
Mid-Point of Raw Response Interval

